"Some familiarity with command line Unix/Linux is a basic research skill."

Is that true? Why?

"As of November 2014, 97% of the world's 500 fastest supercomputers run some variant of Linux, including the top 80."

(`http://en.wikipedia.org/wiki/Linux`)

In 2012, 90%, including the 10 fastest.

Take a look over the tutorials at `http://www.ee.surrey.ac.uk/Teaching/Unix`. Does the material in Tutorials 1–5 seem fairly familiar, or entirely unfamiliar?

entirely or almost entirely unfamiliar: 40%

somewhat familiar: 40%

confident user: 20%

"I have made the switch from Windows to Linux 2 years ago and will probably not go back."

AS PREPARATION FOR A RESEARCH CAREER IN STATISTICS, I ENCOURAGE THIS SWITCH!

Why Linux?

Do statisticians even need to be able to use powerful scientific computing environments, such as clusters?

Why R? What are the alternatives? How about Matlab, or Python, or something else...

not much R experience: 10%

moderate R confidence or more: 90%

R topics of particular interest: packages, object oriented programming

R recommendations: Hadley Wickham's books; the plyr package

Often it is convenient to link some simple C code with R to make the most computationally intensive part of an analysis run quickly. How much programming have you done in a low-level language like C or C++?

essentially none: 30%

one or two classes, not much else: 50%

confident user: 20%

Does the sample C code on the course
website (`http://www.dept.lsa.umich.edu/`
`~ionides/810/gompertz.c`) seem readable to
you? Are you comfortable with the use of
pointers (passing by reference versus
passing by value) in this code?

Googling "C pointers wikipedia" gives

`http://en.wikipedia.org/wiki/Pointer_`
`computer_programming`

(This code uses the `Rmath.h` header file to access versions of various R functions from within C. This is described at `http://cran.r-project.org/doc/manuals/R-exts.html`)

What is Flux? (google: "flux umich")

Only one response indicated regular use of git. Two responses indicated some use. Many others expressed interest in learning this tool.

Who uses git/github and why?

Should we be using it?

A bit of history:

Git was developed by Linus Torvalds in 2005, when he decided no existing code management system was suitable for development of the Linux kernel.

(`https://en.wikipedia.org/wiki/Git_%28software%29`)

Git provides the foundation for GitHub, the largest source code repository in the world. Bitbucket is another notable git repository, among many others. It is also fairly easy to use git to share files privately, without using an internet repository.

A powerful statistical research environment

Linux operating system

Latex document preparation

R data analysis and software development

C/C++ only for time-limited parts of the computation

Linux clusters for computationally intensive analysis

R packages for software communication

git for development and sharing of code

What do all these components have in common?